

# PATENT ABSTRACTS OF JAPAN

(11)Publication number : 11-045097

(43)Date of publication of application : 16.02.1999

(51)Int.Cl.

G10L 3/00

G10L 5/06

(21)Application number : 09-201685

(71)Applicant : NEC CORP

(22)Date of filing : 28.07.1997

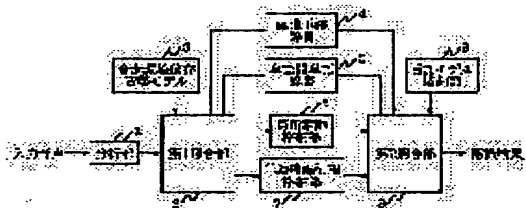
(72)Inventor : ISOTANI RYOSUKE

## (54) CONTINUOUS VOICE RECOGNITION SYSTEM

### (57)Abstract

**PROBLEM TO BE SOLVED:** To perform large vocabulary continuous voice recognizing at high speed using a phoneme environment depending acoustic model.

**SOLUTION:** A recognition word dictionary 4 describes an acoustic model group decided without depending on words of before and after about each word of a recognition object vocabulary as a recognition word. An inter-word dictionary 5 describes an acoustic model group used depending on words of before and after in a word border as inter-word. A first collating section 2 collates a time series of a feature parameter obtained by analyzing an input voice with the recognition word and the inter-word, and a word corresponding to a score when it is assumed that the group reaches a terminal of a word in each time of a time series of the feature parameter is outputted as word terminal information 7. A second collating section 8 refers to this word terminal information, collates again a time series of the feature parameter with the recognition word and the inter-word, and a word corresponding to a prescribed score is outputted with a mode decided by a system.



## LEGAL STATUS

[Date of request for examination] 28.07.1997

[Date of sending the examiner's decision of rejection]

[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]

[Date of final disposal for application]

[Patent number] 3042455

[Date of registration] 10.03.2000

[Number of appeal against examiner's decision of rejection]

[Date of requesting appeal against examiner's decision of rejection]

[Date of extinction of right] 10.03.2003

(11) 特許出願公開番号

特開平11-45097

(5)Int.Cl.	識別記号	FI	請求項の数	OL (全6頁)
G10L 3/00	531	G10L 3/00	531D	
5/06		5/06	D	

(21)出願番号	特願平9-201685
(22)出願日	平成9年(1997)7月28日

(71) 出題人 00000-4237 日本電氣株式会社  
東京都港区芝五丁目7番1号

(72) 説明者 鎌谷 栄輔 日本電機株式会社  
東京都港区芝五丁目7番1号

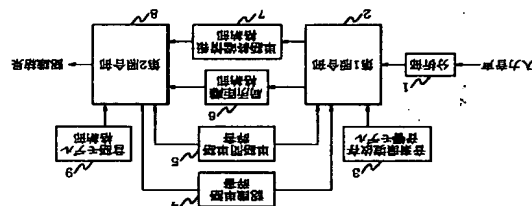
(74) 代理人 糸田士 京本 直樹 (外2名)

## (54)【発明の名称】連続音声認識方式

57)【要約】音楽環境依存音響モデルを用いた大規模楽譜処理方法の提案を行う。

【課題】音楽環境依存音響モデルを用いた大規模楽譜処理方法の提案を行う。

【解決手段】楽譜単部番号4は認識対象楽器中の各単部番号について前後の単語に依存せずに決まる音響モデル系列を登録単語として記述する。単語間単語番号5は単語境界上において前後の単語に依存して用いられる音響モデル系列を単語間単語として記述する。第1の照合部2は、入力音響データを分析して得られた特徴パラメータの時系列データの時系列の各時刻において単語の終端に到達したとき、単語の時系列のスコアと対応する単語を単語候補情報7として出力する。第2の照合部8はこの単語候補情報7を照して再度特徴パラメータの時系列を前記認識対象楽器の単語候補と照合し予め定められた形態で出力する。



(3)

【0007】この問題に対し、単語内には音素環境依存の音素モデルを用い、単語境界では環境に依存しない音素モデルを使用する連続音声認識方式が、特開平5-24692に開示されている。これにより、単語間の処理量の増大を抑えることができる。これを前述の4単語の例に適用すると、図8のようになる。ここで、たとえば、「(o) (\*)」は後続音素によりなる $o/o/$ の音素モデルを表す。全体として照合すべき音素モデルの数は16に削減される。大略案の場合は、さらに削減効果は大きい。

【0008】  
【発明が解決しようとする課題】 上述した従来の連続音声認識方式は、精度を向上させるために音素環境依存モデルを使用すると、単語境界での処理量が增大するという問題がある。これに対し、単語内には音素環境依存の音素モデルを用い、単語境界では環境に依存しない音素モデルを使用すると、単語境界での処理量の増大を抑えることができるが、一方で単語境界に用いる音素モデルの精度が低いために、とくに大略案の連続音声認識では性能の低下を生じるおそれがある。

【0009】本発明の目的は、単語境界にも音素環境依存の音素モデルを用いて精度を確保しつつ、大略案の場合でも単語境界での処理量の増大を抑えることのできる連続音声認識方法を提供することにある。

【0010】  
【課題を解決するための手段】 本第1の発明の連続音声認識方式は、音素環境依存音素モデルを用いる連続音声認識方式において、認識対象音素中の各単語について前後の単語に依存せず決まる音素モデル系列を認識単語として記述した認識単語と、単語境界に於いて前後の単語に依存して用いられる音素モデル系列を単語間単語として記述した単語間単語と、入力音声を分析して特徴パラメータの時系列を得る分析部と、前記特徴パラメータの時系列を前記認識単語および前記単語間単語と照合して該特徴パラメータの時系列の各時刻にて単語の終端に到達した仮説のスコアと対応する単語を単語候補情報として出力する第1の照合部と、前記単語候補情報を参照して再度前記特徴パラメータの時系列を前記認識単語および前記単語間単語と照合しスコアに基づいて単語の候補をシステムで定められた形態で出力する第2の照合部を有することを特徴とする。

【0011】また、第2の発明は、第1の発明における前記第1の照合部が前記各時刻において前記スコアがあらかじめ定められた基準を満たさない仮説については照合を打ち切り、前記単語候補情報と対応する各時刻において少なくとも単語終端に到達した仮説に於いて認識単語および単語間単語を出力することを特徴とする。

【0012】また、第3の発明は、第1の発明および第2の発明における前記第1の照合部で各時刻ごとに計算される局所距離を格納する局所距離格納部を備え、前記

(4)

の際、音素モデル格納部9に記述された単語間の接続情報などを参照して用いる。  
【0020】音素モデルとしては、たとえば第一照合部ラムモデルを用いることができる。一般には第一照合部と第二照合部で異なる音素モデルや異なる特徴パラメータを用いることもできるが、同じ音素モデルと特徴パラメータを用いる場合には、第一照合部で計算し局所距離格納部6に保存する場合に、局所距離の値を用いることができ、第二照合部では、アクティブな単語（単語間単語を含む）の始端に対して先行する単語のうち第一照合部で終端がアクティブであった単語についてのみ仮説を展開する。また、第一照合部で終端がアクティブであったも、その累積スコアと、第二照合部での時点までの累積スコアの値に比べて、見込みの小さい単語については仮説を展開しないようにすることもできる。

【0021】以下、認識単語格納部4と単語間単語格納部5について、従来の技術の説明に用いた簡略化した例に基づいて、図2を参照して説明する。この例では、音素モデルとして後続の音素に依存する音素モデルを使用しており、辞書は「ここ」「そこ」「から」「まで」の4単語からなる。たとえば、「そこ」は、 $s/o/$ 、 $/o/$ 、 $/o/$ 、 $/o/$ の4音素からなり、前の3つの音素に対応する音素モデルはそれぞれ「s(o)」、「o」、「(k)」、「k(o)」と決まると、最後の $o/o/$ については、後述に「から」がくる場合は「o(k)」、「(k)」、「(k)」、「(k)」となるなど、一意的に決まらない。そこで、認識単語格納部では、単語「そこ」の音素モデル系列として「s(o)」、「o(k)」、「(k)」、「k(o)」のみを記述し、「o(k)」、「(k)」、「(k)」、「(k)」を省略する。

【0022】図2の例に対応した認識単語格納部4および単語間単語格納部5の構成例を、それぞれ図3、図4に示す。単語辞書には、認識対象音素中の各単語に対し、表記、その単語を構成する音素モデル系列の情報に加え、始端、終端のカテゴリが格納されている。単語間辞書には、先行単語の終端カテゴリと後続単語の始端カテゴリの組合せに対し、それらの単語間に用いられる音素モデル系列の情報が記述されている。

【0023】図2では全体として照合すべき音素モデルの数は21で、図7に示した従来の例の場合の24に比べて削減されている。この例は4単語からなる簡略化した例であるため効果はそれほど大きくないが、大略案の場合には後続音素環境の複雑がふえるため、削減効果は大きい。

い、図8に示した従来例に比べて照合すべき音素モデルの数は増えているが、単語間で図2の例と同様に音素環境依存音素モデルを使用しているために、精度の高い照合が可能となっている。

【0024】音素環境としては、音素そのものでなく、いくつかの音素をまとめた音素クラスを用いることもできる。また、後続音素にのみ依存する音素モデルのかわりに、前後の音素に依存する音素モデルを用いることもできる。その場合は、認識単語辞書には最初の音素と最後の音素を除いた音素モデルを記述し、後続音素に依存する単語終端の音素と先行音素に依存する単語始端の音素を除いた音素モデルは、お互いに接続し合うものを組にして単語間単語辞書に記述しておく。また、前後の音素に依存する音素モデルを用いる場合は、単語辞書および単語間単語辞書の記述例を、それぞれ図5、図6に示す。図で、たとえば「(s) o (k)」は、先行音素が $s/$ で後続音素が $k/$ である音素 $o/o/$ の音素モデルを表す。

【0025】図2の例では、各単語の音素モデル系列を独立に扱っているが、一般に大略案の場合は先頭部分の共通な単語が多く、全単語の音素モデル系列を先頭を共通化した共通部分の形で表現することも可能である。

【0026】  
【発明の効果】 以上説明したように、本発明によれば大略案の連続音声認識と、音素環境依存音素モデルを用いて高精度に、しかも従来の音素環境依存音素モデルを用いた場合より処理量を削減して行うことができる効果がある。

【図面の簡単な説明】  
【図1】本発明の実施の一形態を示すブロック図である。

【図2】図1の第1照合部で構成される単語のネットワークの例を示す図である。

【図3】本発明の実施例における、認識単語辞書の例を示す図である。

【図4】本発明の実施例における、単語間単語辞書の例を示す図である。

【図5】本発明の実施例における、認識単語辞書の別の例を示す図である。

【図6】本発明の実施例における、単語間単語辞書の別の例を示す図である。

【図7】従来の例における、単語終端の音素モデルの例を示す図である。

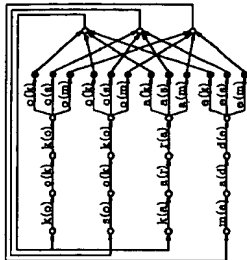
【図8】従来の例における、別の単語終端の音素モデルの例を示す図である。

【符号の説明】

- |   |             |
|---|-------------|
| 1 | 分析部         |
| 2 | 第1照合部       |
| 3 | 音素環境依存音素モデル |
| 4 | 認識単語辞書      |

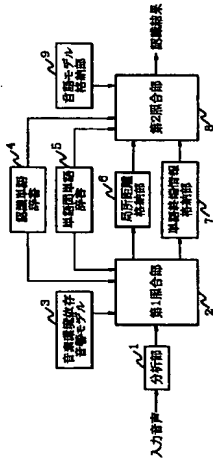
(6)

【図7】

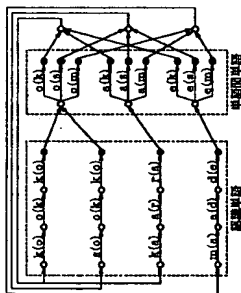


(5)

【図1】



【図2】



【図3】

番号	表記	言語モデル列	状態	結果
1	二	k(o) o(k) k(o)	k	o
2	二	k(o) o(k) k(o)	o	o
3	三	k(a) s(r) r(a)	k	a
4	三	m(a) s(d) d(a)	m	a

【図4】

番号	先行単語列	後置単語列	言語モデル列
1	o	k	o(k)
2	o	a	o(a)
3	o	m	o(m)
4	a	k	a(k)
5	a	a	a(a)
6	a	m	a(m)
7	a	k	a(k)
8	o	a	o(a)
9	o	m	o(m)

【図5】

番号	表記	言語モデル列	状態	結果
1	二	k(o) o(k) k(o)	ko	ko
2	二	k(o) o(k) k(o)	ao	ko
3	三	k(a) s(r) r(a)	ka	ra
4	三	m(a) s(d) d(a)	ma	da

【図6】

番号	先行単語列	後置単語列	言語モデル列
1	ko	ko	k(o)k(o) k(k)o
2	ao	ao	k(o)a(o) a(a)o
3	ko	ka	k(o)k(a) k(a)k(a)
4	ko	ma	k(o)k(a) m(a)k(a)
5	ra	ko	r(a)k(o) k(o)k(a)
6	ra	ao	r(a)k(a) a(o)k(a)
7	ra	ka	r(a)k(a) k(a)k(a)
8	ra	ma	r(a)k(a) m(a)k(a)
9	da	ko	d(a)k(o) k(o)k(a)
10	da	ao	d(a)k(a) a(o)k(a)
11	da	ka	d(a)k(a) k(a)k(a)
12	da	ma	d(a)k(a) m(a)k(a)

【図8】

